# Introduction to text mining Twitter with R

Paul Nulty

July 2, 2014

# List of Acronyms and Definitions

- API: Application Programming Interface — a specification of how computer programs can communicate with each other

- JSON: JavaScript Object Notation — a human readable format for representing attribute-value pairs.

- OAuth: An open standard for authorization. OAuth provides client applications a 'secure delegated access' to server resources on behalf of a resource owner.

- HTTP: HyperText Transfer Protocol: Protocol specification for requesting and sending hypertext over the internet.

- REST: REpresentational State Transfer: A type of communication between software over the internet by sending and receiving data with requests and responses, usually with the request sent as a URL and the response in XML or JSON.

# Useful Links

`https://dev.twitter.com/docs/things-every-developer-should-know`

```
2+3

## [1] 5

library(quanteda)
tokenize("Just testing sweave/knitr code integration here.")

## [1] "just"        "testing"     "sweaveknitr" "code"
## [5] "integration" "here"
```

# The twitter APIs

Twitter offers two ways for computer programs to post and retrieve data from its service. The twitter REST Search API will receive requests as http commands, and will respond with a JSON object representing a list of tweets. The twitter streaming API allows a computer program to maintain a connection to twitter's service and receive a flow of live public tweets that match the filters or search terms specified.

Both of these APIs require that the developer of the computer program create a twitter account, create a new application, and receive an access token.

# Getting programmatic access to twitter

[1]

1. Go to https://dev.twitter.com/ and sign in with a twitter account

2. Go to My Applications and select "New App". https://apps.twitter.com/

3. Fill out the form and click 'create'

4. The next page will ask for application details. The first three are required so you will need to create a Name, Description, and Website URL. You can choose whatever you want for these fields but keep in mind that the name can not contain the word "Twitter". You should leave the "Callback URL" blank.

5. You should now see page a page that says "Application Management" on the top, with the name of your app. Click on the API Keys tab.

6. You should now see a list of "Application settings". The first two here are the most important and you will need to copy and paste these into a file. The first is the API key. The second is the API secret. Both of them should be long strinigs of garbage looking letters and numbers. These are

---

[1]This section is adapted from `https://openhatch.org/wiki/Community_Data_Science_Workshops/Twitter_authentication_setupandiscorrectasofMay1st2014`.

the keys that will let you program know that it's approved to grab data from Twitter.

7. At the bottom of the page you should see a section marked "Your access token" that says, "You haven't authorized this application for your own account yet." Below that there is a button that says "Create my access token". Click on that button. Once you press it there will be a yellow 'Status' box at the top. It will give you a link you can click to refresh. Wait a minute or two and click it untill you see information, below including lines that give you your "Access token" and your "Access token secret". Copy these two pieces of information.

8. You should have four long strings of letters and numbers taken from this page now:API key, API secret, Access token, Access token secret. These are the keys requested by most programmatic twitter interfaces.

## Using the REST API

Accessing twitter data through the REST API works by constructing a HTTP GET request — the same kind of request a web browser sends when requesting a web page. However, rather than responding with HTML, the twitter API will respond with a JSON representation of tweet data. We can construct and send HTTP requests in R using the RCurl package, and parse the response with rjson, but there is an R package which simplifies this process, acting as an R client to the twitter service: twitteR.

After installing twitter, the first step is to register your API keys for your application using OAuth: The four strings

library('twitteR') $setup_t witter_o auth('ZOHJIRAwnw23FhvFWyUg',' HTfEcEmRRDcx0ZsJ5CHOcm$ $ZkDTfl3IbIFZFXlVokA6Gpc19TZPyov3wucZ0XaB',' 8vgPnpEWP3qhvILmTLXVb9RslwcEwVVeKO$